DOCUMENT RESUME

ED 142 128 HE 009 116

AUTHOR Wood, Peter H.
TITLE The Description and Evaluation of a College Department's Faculty Rating System.
PUB DATE 77
NOTE 26p.; Paper presented at the annual meeting of the American Educational Research Association (New York, April 1977) ; Page 26 may be marginally legible due to print quality

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *College Faculty; *Departments; Experiments; *Faculty Evaluation; Higher Education; Offices (Facilities); *Peer Evaluation; Rating Scales; *Student Evaluation of Teacher Performance; Systems Approach
IDENTIFIERS Department Chairpersons

ABSTRACT

Faculty in a medium-sized (20 to 30 member) college department were rated as teachers by their students, departmental colleagues, and departmental chairperson. Departmental faculty were also rated as researchers by their departmental colleagues. Several ratings and ranking techniques were employed over a three-year period. Colleague ratings of research and professional service, and student ratings of teaching produced ratings of higher reliability and consistency than did colleague ratings of teaching. Colleague ratings of teaching were more strongly influenced by office location than were ratings of research and service. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED142128

THE DESCRIPTION AND EVALUATION OF A COLLEGE DEPARTMENT'S FACULTY RATING SYSTEM

Peter H. Wood
C/O Educ. Foundations and Inquiry
Bowling Green State University
Bowling Green, Ohio 43403

Objectives of the Study

This study was undertaken to assess the reliabilities and inter-instrument correlations that characterize various rating procedures used to evaluate the faculty of a medium-sized college department. At Bowling Green State University, the department chairperson is obligated to assess three dimensions of faculty performance: 1) Teaching; 2) Research and/or Scholarly Activities; and 3) Service. Results of these evaluations are used to determine: 1) Reappointment; 2) Tenure; 3) Promotion; 4) Salary Increases; and to some degree, 5) Teacher Assignment. Several evaluation procedures have been employed during the past three years. This study represents an analysis of some of the characteristics of these procedures.

Perspective--Theoretical Framework

Reduced rates of college expansion and demands for fair personnel procedures have caused college administrators to examine their faculty evaluation procedures and to make them more objective. In the past several years, department chairpersons at BGSU have: been sued by faculty claiming unfair hiring, retention, or salary policies; been forced to reallocate faculty lines due to changes in student enrollment; and been asked by the

HE010 116

Board of Trustees to institute a differential, merit-based system of salary increases. Four years ago, the Educational Foundations (EDFI) Department at BGSU established a series of committees to investigate faculty evaluation procedures.

The majority of the departmental procedures that were observed seemed to be categorizable into one of three types:

I. Non-Empirical -- Administrators and/or selected faculty members meet in committees to examine vita and make whatever personnel decisions are required.

II. Empirical-Ratings -- Students, peers, and/or others are asked to rate faculty performance on a set of common scales, and these ratings are somehow combined with committee or administrative opinions to produce personnel decisions.

III. Criterion-Referenced -- Specific performance criteria are established for individual faculty, and faculty are evaluated according to the degree to which they meet these criteria.

This study is a presentation of some of the results produced by the ratings procedures characteristic of the Type II (Empirical-Ratings) approach to faculty evaluation.

Context and Some Limitations of the Study

Personnel evaluation in a collegial setting--especially in an institution which faces a potential reduction and/or reallocation of staff--presents a wide range of problems. Each evaluation effort threatens some of those who are asked to support or contribute to it.[1] While increased experience with evaluative procedures and increased pressures to produce an objective system create a movement toward a more criterion referenced system,

general faculty resistance to evaluation for personnel-decision purposes creates a counter force toward a more casual and less objective approach. Much of the data that is missing from this report is missing because: 1) each year the majority of the departmental faculty were supportive of somewhat different procedures; 2) individual faculty failed to participate in the generally supported procedures because of various logical or ethical considerations; and 3) some data--especially information related to the rating responses of individual faculty--were intentionally obliterated to protect the anonymity of the raters.

The size of the department is small--especially in comparison to the number of hypotheses which could be generated concerning faculty perceptions and activities. The rating instruments and procedures fall far short of perfection since they were generated more to reflect the shifting consensus of departmental faculty than to reflect the current state of the psychometric art. Insofar as these measures reflect an Empirical-Ratings stage of personnel procedure, one could best describe them as "early" or "general" ratings. With additional experience with ratings use, it is possible that there may be a shift to more behaviorally-defined ratings scales and option keys--provided that there is a common agreement as to those behaviors which represent various degrees of teaching, research or service performance.

Instruments, Data Sources and Assessment Techniques

The three primary faculty functions of 1) Teaching, 2) Research/ Scholarship, and 3) Service cause faculty to interact with different potential raters of these functions.

Student Ratings (1974, 1975, 1976): A student rating form[2] was developed and modified over a two-year period. The resulting form caused students to rate faculty on several dimensions (scholarship, organization/clarity, interaction with group, interaction with individuals, and enthusiasm)[3] to orient them to characteristics often valued in a teacher. It then asked students to produce a general assessment of: 1) the teacher; 2) the total course experience; and 3) their own accomplishment in the course. The results of the three general questions were averaged to produce the student rating score. Twenty EDFI faculty were rated by their students at the end of the Winter Quarter, 1974. All EDFI faculty used the same form to produce the student ratings scores for 1975 and 1976. The score for 1975 was the total mean score resulting from all student responses to the three key questions for three separate quarters, Spring 1974, Fall 1974 and Winter 1975. The 1976 scores were similarily created.

Peer Ratings (1974, 1975, 1976): All faculty of the EDFI Department were asked to rate all other faculty on the three dimensions of teaching, research, and service. The 1974 form consisted of three five-point scales in which the one position was defined as "low" and the five position was defined as "high." The 1975 and 1976 peer forms asked all faculty to rank department members from first up to seventh on each of the three dimensions. The form listed several criteria which were considered to be relevant to each dimension. All peers were given access to all department personnel files which contained vita, letters of recommendation and other data. Non-ranked faculty were automatically assigned the ranking of eight. This ranking procedure resulted from faculty complaints that they could not honestly rank or rate all department members since they were unknowledgeable concerning

the activities of many. The 1975-76 peer form produced two statistics for each faculty member on each dimension--the number of times ranked in the top seven, and the total ranking score (with non-rankings equal to eight).

Chairperson Ratings (1974): In 1974 the chairperson rated all faculty on the three dimension, five-point scale used by peers. There was no independent chairperson ranking or rating in 1975 or in 1976.

Committee Ratings (1975): In 1974, a faculty evaluation committee was created--one member elected from each of the four departmental ranks (instructor, assistant professor, associate professor, full professor), and the fifth person chosen by the department chairperson so as to cause both sexes and all departmental sub-divisions to be represented on the committee. In 1975, this committee independently examined the vita of all faculty and rated each on five-point scales for teaching, research, and service. The five positions on each scale were labeled, and several lead-in questions were used to orient the committee members to criteria believed to be relevant to each faculty function. There was no committee rating in 1976.

Visibility: Each faculty member was categorized as to visability to other faculty. Faculty with offices adjacent to the departmental office were labeled as highly visible 1; faculty with offices on the two main corridors near the departmental office were labeled 2 for the central corridor and 3 for the next most central corridor. Faculty in the rear corridor were labeled 4; faculty on a different floor of the building were labeled 5; and faculty with offices in another building were labeled as 6 (lease visible to other departmental members).

Rank: Faculty were also categorized as to their faculty rank at the beginning of each of the three years.

Area: Faculty were also characterized as belonging to one of four

subdivisions existing within the department.

College Personnel File: Each Spring, every department chair is required to file a "Substantiation for Salary and Promotion Recommendations" form which presents the salary, contract type, rank and effectiveness rating for each faculty member of the department. All faculty members were rated by the chair as to their Teaching, Research-Service-Scholarship, and University Service. The labels for the five-point scale used for this form are:

1 = Outstanding

2 = Superior

3 = Above Average

4 = Average

5 = Below Average.

These ratings were available for the 1973-74 year, the 1974-75 year, and the 1975-76 year. Salaries were available for these same years. The three ratings for each year were created by the chairperson who reorganized the various peer and student ratings through use of formulas which shifted each year according to faculty or evaluation committee decisions.

Analysis of Data

The various ratings and faculty categories were compared through use of bivariate correlation analyses--Pearson product-moment correlation, Spearman's rank order correlation and Kendall's rank order correlation (used when there were many tied ranks). Ratings procedures were analyzed for reliability via analysis of variance. Ratings and area, rank and visibility identifications were examined via an analysis of variance with area or rank or visibility identification functioning as the independent variables.

Results

College Personnel File ratings were created by the department chair for each faculty member for the academic years of 1973-74, 1974-75, 1975-76. Although the formulas used to produce these ratings varied from year to year, each was created primarily from some combination of peer and student ratings. The student ratings were blended with peer ratings to produce the Teaching scores but not the Research/Scholarship or University Service scores. The pattern of Pearson product-moment correlation coefficients seems to indicate that: 1) the Teacher ratings are relatively consistent across the three years--as are most of the Research/Scholarship and Service ratings; 2) the Teacher ratings are generally unrelated to the Research/Scholarship and Service ratings; but 3) the Research/Scholarship and Service ratings are quite closely associated with each other.

TABLE 1

CORRELATIONS BETWEEN YEARLY RATINGS OF THREE FACULTY FUNCTIONS: TEACHING (T), RESEARCH/SCHOLARSHIP (R), SERVICE (S), (N = 20 to 24)[3]

		1	2	3	4	5	6	7	8
T	1. 1974	--							
T	2. 1975	62	--						
T	3. 1976	68	72	--					
R	4. 1974	08*	26	07	--				
R	5. 1975	-08	37*	12	69	--			
R	6. 1976	-12	15	24*	27	62	--		
S	7. 1974	08*	-05	05	64*	55	39	--	
S	8. 1975	-06	34*	20	62	74*	50	55	--
S	9. 1976	09	50	39*	48	29	09*	25	69

Notes: Correlation coefficient decimal points have been removed.
Underlined coefficients reflect a common function across years.

*These correlations reflect a common year but not a common function.

The Evaluation of the Teaching Function produced the most varied types of ratings. In the 1973-74 year, teaching was rated by 18 of the 24 members of the department; by students (for only one term, Winter 1974); and by the department chair. During the following year (1974-75): most peers indicated their rankings (from one to seven) of the best teachers; the five members of the evaluation committee rated all faculty on a five point scale; and three quarters of student ratings were added to the pool. The same (1974-75) peer ratings and student ratings of teaching were again employed in the 1975-76 academic year.

TABLE 2

CORRELATIONS BETWEEN PEER, CHAIR, EVALUATION COMMITTEE, AND STUDENT RATINGS OF TEACHING: 1974 to 1976. (N = 19 to 26)

	1	2	3	4	5	6	7
1. Student '74	--						
2. Student '75	30	--					
3. Student '76	46	49	--				
4. Peer '74	64	13	44	--			
5. Peer '75	39	-12	13	67	--		
6. Peer '76	35	18	33	43	76	--	
7. Chair '74	40	30	23	63	55	37	--
8. Committee '75	38	09	44	70	63	60	68

Notes: Correlation coefficient decimal points have been removed.
Underlined coefficients reflect a common type of rater.
All scales have been converted to reflect a similar direction.

The consistency of student ratings across the three years was not very impressive. The peer ratings of teaching appear to be somewhat more consistent across the three years even though different procedures and forms were used to elicit them during the first year. The committee-peer, committee-chair

and peer-chair ratings are similar to the other peer ratings. The relationships between peer and student ratings for two of the three years is similar in nature to the consistency of student ratings across years--hardly impressive.

Some of the low correlations presented in Tables 1 and 2 may be partly attributed to a real inconsistency in the beginning performances of faculty new to the department--or inconsistency in the way that other faculty perceive their performance.[4] When the data associated with the newer faculty--those entering the department after 1971--are removed from the analyses, two new tables are created.

TABLE 3

CORRELATIONS BETWEEN YEARLY RATINGS OF THREE FACULTY FUNCTIONS: TEACHING (T), RESEARCH/SCHOLARSHIP (R), SERVICE (S) OF PRE-1972 FACULTY (N=17 to 20)

		1	2	3	4	5	6	7	8
T	1. 1974	--							
T	2. 1975	69	--						
T	3. 1976	69	68	--					
R	4. 1974	24*	20	15	--				
R	5. 1975	-05	31*	05	65	--			
R	6. 1976	-23	-13	-12*	40	65	--		
S	7. 1974	02*	-06	05	61*	55	53	--	
S	8. 1975	02	28*	07	52	75*	44	52	--
S	9. 1976	13	43	29*	39	25	-05*	23	71

Notes: Correlation coefficient decimal points have been removed.
Underlined correlations reflect a common function across years.

*These correlations reflect a common year but not a common function.

TABLE 4

CORRELATIONS BETWEEN PEER, CHAIR, EVALUATION COMMITTEE, AND STUDENT RATINGS OF TEACHING: OF PRE-1972 FACULTY (N=16 to 20)

	1	2	3	4	5	6	7
1. Student '74	--						
2. Student '75	65	--					
3. Student '76	64	71	--				
4. Peer '74	76	27	39	--			
5. Peer '75	53	02	-03	75	--		
6. Peer '76	60	36	17	48	74	--	
7. Chair '74	60	35	18	72	72	54	--
8. Committee '75	61	20	40	68	69	61	67

Notes: Correlation coefficient decimal points have been removed
Underlined coefficients reflect a common type of rater.
All scales have been converted to reflect a similar direction.

Comparisons between Tables 1 and 3 seem to indicate that eliminating the data from newer faculty has little effect upon the resulting correlations among the college ratings. A similar comparison between Tables 2 and 4 does seem to indicate some change. The three-year consistency of the student ratings improves as does the apparent relationship between the student ratings of the first year the department used a common form (1974) and the various peer/committee/chair ratings on that and subsequent years. The Peer Ratings of the three major faculty functions are presented in Table 5.

Some tentative conclusions could be developed from the correlation patterns presented in Table 5: 1) the size of the correlations between the same year's ratings/rankings of the three different faculty functions may be decreasing as faculty gain experience with identifying and evaluating

TABLE 5

CORRELATIONS BETWEEN PEER RATINGS OF TEACHING (T), RESEARCH/SCHOLARSHIP (R), AND SERVICE (S): 1974 TO 1976 (N = 23 to 26)

	1	2	3	4	5	6	7	8	9	10	11
Peer 1 T '74	--										
Peer 2 R '74	56(54)	--									
Peer 3 S '74	44(43)	78(77)	--								
Peer 4 T '75	70(75)	26(34)	03(11)	--							
Comm 5 T '75	70(68)	52(48)	30(28)	66(68)	--						
Peer 6 R '75	30(30)	52(52)	65(65)	42(43)	37(36)	--					
Comm 7 R '75	36(32)	68(64)	71(70)	35(36)	54(49)	78(78)	--				
Peer 8 S '75	18(18)	62(62)	35(35)	34(36)	28(28)	47(43)	40(36)	--			
Comm 9 S '75	26(21)	75(72)	54(49)	28(32)	48(43)	60(57)	65(61)	81(84)	--		
Peer 10 T '76	43(48)	01(04)	27(27)	78(74)	60(61)	13(12)	10(05)	06(08)	08(09)	--	
Peer 11 R '76	-07(-23)	20(26)	35(53)	18(24)	31(08)	39(39)	49(54)	01(05)	16(11)	32(31)	--
Peer 12 S '76	04(01)	39(38)	07(05)	36(32)	28(22)	32(25)	30(23)	84(84)	73(74)	23(22)	02(13)

Notes: Correlation decimal points have been removed.
Underlined coefficients reflect a common rating function.
All scales have been converted to reflect a similar direction.
Coefficients in parentheses represent analyses with data from post-1972 faculty eliminated.

evidences of these functions; 2) the Research/Scholarship and the Service functions were not clearly differentiated by peers during the first two years of peer evaluation; 3) removing the data reflecting the new faculty seemed to have little effect upon the peer-rating/ranking correlation matrix; and 4) correlations are slight for any one function across more than one year. Any of these possible trends would have to persist for several more years before they could be described as being more than heuristic hypotheses.

There are two ways of obtaining peer rankings when you ask peers to rank the best seven departmental members as to effectiveness. The peer rankings in Tables 2, 4 and 5 were obtained by adding up all of the rankings for each faculty member. If a faculty member was not ranked in the top seven faculty members--or was not ranked because another faculty member did not participate in the ranking procedure--a ranking of eight was assigned to the faculty member being evaluated. This ranking of eight was added to the other rankings--if any. Consequently colleagues not ranked by anyone (even by themselves) were credited with a peer ranking score of 208 (26 faculty members in 1975 times 8 = 208). The lowest (best) ranking one could achieve was a value of 26--if all 26 faculty ranked you as first. The actual range of scores was:

1975 Peer Teaching: 151 to 208, mean = 189.3, s.d. = 18.0

1975 Peer Research: 131 to 208, mean = 189.5, s.d. = 18.5

1975 Peer Service: 119 to 208, Mean = 189.8, s.d. = 21.9

1976 Peer Teaching: 156 to 240, mean = 218.9, s.d. = 19.1

1976 Peer Research: 175 to 240, mean = 219.3, s.d. = 15.6

1976 Peer Service: 135 to 240, mean = 218.2, s.d. = 22.0

As is obvious from the scores, while 26 faculty were members in the department for the Spring ranking of 1975, thirty faculty could have

participated in the next year's ranking (1976). Due to a committee decision to maintain anonymity of the rankers, the 1976 data was destroyed as soon as rank sums were created. Therefore, there is no additional data for the 1976 peer rankings.

The other effectiveness measure created by this ranking procedure is the number of times a colleague was ranked in the top seven for one of the three faculty functions. The Pearson Product-Moment correlations between these two numbers--a sum ranking with nonrankings equal to eight, and the number of times ranked--was -.96, -.98, and -.94 respectively for Teaching, Research/Scholarship, and Service. In 1975, twenty-one of the twenty-six faculty were ranked in the top seven as teachers, and twenty-three of twenty-six were ranked for Research/Scholarship, and for Service.

Peer Ratings may reflect bias of various sorts. The department is subdivided into four separate areas, and area identification may influence ratings. Different faculty joined the department at different times, and groups entering during similar periods may form cohorts which influence peer ratings. Office locations may influence peer-interaction and so influence peer ratings. Unfortunately, peer ratings are anonymous, and the rater characteristics are not available for investigation. However, the characteristics of the rated peers may be compared with their ratings. Any discovered relationships may reflect bias--or they may reflect a reasonable and logical relationship with performance levels. Table 6 presents some of the relationships between ratings or rankings and area, year joined department, and office location.

TABLE 6

RELATIONSHIPS BETWEEN STUDENT (S), PEER (P) AND COMMITTEE (C) RATINGS AND AREA, YEAR ENTERING DEPARTMENT, AND OFFICE LOCATION

Rating Source	AREA			YEAR				OFFICE			
	F	(df)	Eta^2	F	(df)	Eta^2	r	F	(df)	Eta^2	r
P Teaching '74	1.25	(3,18)	.17	1.75	(13,9)	.71	-.03	2.40	(5,17)	.41	.44**
P Research '74	.52	(3,18)	.08	.62	(13,9)	.48	-.20	1.42	(5,17)	.29	.10
P Service '74	2.0	(3,18)	.25	.72	(13,9)	.51	-.08	1.14	(5,17)	.25	.19
S Teaching '74	.33	(3,14)	.07	1.19	(10,8)	.60	-.22	.68	(4,14)	.16	.09
P Teaching '75	1.71	(3,21)	.20	2.96	(14,11)	.79	-.05	10.46	(5,20)	.72	.58***
C Teaching '75	.53	(3,21)	.07	1.37	(14,11)	.64	.02	3.39	(5,20)	.46	.37
P Research '75	.70	(3,21)	.09	.88	(14,11)	.53	.11	2.56	(5,20)	.39	.37**
C Research '75	.98	(3,21)	.12	1.29	(14,11)	.62	.08	1.78	(5,20)	.31	.29
P Service '75	.39	(3,21)	.05	.67	(14,11)	.46	.12	2.36	(5,20)	.37	.34**
C Service '75	.32	(3,21)	.04	.63	(14,11)	.44	.13	2.84	(5,20)	.42	.14
P Teaching '76	1.55	(3,20)	.19	3.89	(13,11)	.82	.00	3.41	(5,19)	.47	.37**
P Research '76	.60	(3,20)	.08	1.24	(13,11)	.59	.06	1.19	(5,19)	.24	.21
P Service '76	.51	(3,20)	.07	.35	(13,11)	.29	.07	.76	(5,19)	.17	.23
S Teaching '76	2.29	(3,21)	.25	.66	(14,10)	.48	-.20	1.00	(5,19)	.21	-.05

Notes: "F" is the ANOVA "F" ratio of mean squares.
"r" is the Kendall correlation coefficient.
All scales have been converted to reflect a similar direction.

**Correlation coefficient is significant beyond the .01 level.
***Correlation coefficient is significant beyond the .001 level.

The data presented in Table 6 can be interpreted as an indication that office location--or visibility to other faculty--might bias peer ratings of teaching in favor of those faculty which have offices in areas which are more centrally located within the distribution of departmental offices. Visibility may also influence peer ratings of research and service. There are faint hints that: 1) the peer ranking approach used in 1975 and 1976 may be more open to visibility bias; and 2) experience with ranking of peers may reduce this "visibility" bias. Student ratings of teaching seem relatively unrelated to office location and to area identification but are slightly related to the year that faculty began teaching in the department--with the more experienced teachers eliciting slightly higher ratings. Some of this relationship may be related to the increased power to teach graduate students or preferred classes that may be gained with longevity within the department.

Data concerning instrument reliabilities is now being collected. Table 7 presents some of the data collected from some of the procedures. The "reliability" figure was derived from a comparison of the mean sum of squares between teachers and the mean sum of squares within each teacher's ratings or rankings. The formula used is: the reliability estimate (r) = (F-1)/F. [5]

The popularity of the departmental evaluation system seemed to be relatively low. The peer rating system used in the spring of 1974 (all faculty rate all other faculty) was voted out in the fall of 1974. The peer ranking system used in 1975 and in 1976 has yet to be voted out of use, but a departmental vote in 1975 caused the separate rating by the elected, five person, Faculty Evaluation Committee to be eliminated. The most recent departmental vote was quite strongly in favor of increasing the participation of the departmental chair in the evaluation of faculty. The same departmental vote

TABLE 7

RELIABILITY ESTIMATES OF VARIOUS DEPARTMENTAL RATING AND RANKING PROCEDURES

	Number of Raters	Anova F Ratio	Estimate of Reliability $(r=(F-1)/F)$[5]
Student Ratings: Fall 1974	1609	13.00	.92
Peer Ratings, '74 Teaching	20	2.61	.62
Peer Ratings, '74 Service	20	2.71	.63
Peer Ratings, '74 Research	20	2.68	.63
Fac. Eval. Committee Ratings, '75 Teaching	5	3.83	.74
Fac. Eval. Committee Ratings, '75 Service	5	6.31	.84
Fac. Eval. Committee Ratings, '75 Research	5	5.08	.80
Peer Rankings, '75 Teaching '75	18	4.64	.78
Peer Rankings, '75 Service '75	18	5.61	.82
Peer Rankings, '75 Research '75	18	5.54	.82

was also in favor of reducing the weight of the peer rankings and of increasing the weight given to student ratings of teaching. For the past four years, the university has requested colleges and departments to provide some sort of evidence that faculty merit was being identified and rewarded at the departmental level. Much of the previously described effort was in partial response to this request. A reduction in university pressure might easily result in an elimination of all peer or student ratings or rankings--at least for tenured faculty who are not within one year of promotion.

Conclusions

When the number of analyses exceed the number of subjects, any conclusions must be regarded with considerable caution. The following conclusions therefore are categorized as: (I) Tentative; and (II) Very Tentative.

I. Tentative conclusions:

1) Student ratings of teaching do not parallel faculty ratings or rankings of teaching--possibly because different criteria are applied by each group;

2) Student ratings of teaching are relatively stable across a three year period--for experienced faculty;

3) Peer ratings or rankings of teaching are also relatively stable--if less so that student ratings--but may be influenced by non-teaching-related variables such as faculty "visibility";

4) Peer ranking systems which permit peers to rank only the "better" faculty are preferred by faculty to any system which requires faculty to rate or rank all faculty of a 20 to 30 person department;

5) Such peer ranking may produce ranking with a consistency (reliability?) at least as good as that characterizing an "all rate all" system;

6) A peer committee may produce ratings which are similar in nature to the rankings produced by an entire department.

II. Some of the more tentative conclusions are:

1) Faculty with little experience in rating or ranking their colleagues may find it difficult to differentiate between the different faculty functions which are broadly labeled as: Teaching; Research/Scholarship; and Service;

2) Rater or ranker ability to differentiate between these different functions may improve with increased experience;

3) The initial publication of student ratings of teaching--or any other indication of effectiveness--may influence faculty evaluations of teaching (or other functions) for several successive years; and

4) The instution of a formal, faculty evaluation system will stimulate many faculty to develop a wide variety of methods by which they can inform other faculty about an incredible variety of previously unheralded activities. This last comment is not supported by the data already present, but is believed to be true by most members of the department.

Notes/References

1. Guion, Hutchinson, Klein, Statz and Wood have just completed a year-long survey of BGSU faculty attitudes toward the evaluation of faculty. The preliminary results seem to indicate that the majority of faculty are generally in opposition to external evaluation of their efforts. A surprising result was that students were preferred to peers and chairpersons as evaluators of teaching performance. This report is as yet unpublished, but will be submitted to ERIC in the near future.

2. The student rating form used in this study is presented in the appendix of this report. The first five questions were adapted from the general factor titles developed by Hildebrand, Wilson and Dienst and reported in their Evaluating University Teaching (Center for Research and Development in Higher Education, University of California, Berkeley, 52 pages, undated). The questions actually used to produce an evaluation of the teacher's classroom effectiveness are the three very general, judgemental questions which follow the first five Berkeley-derived, orienting questions.

3. In general, those correlation coefficients larger than .39 tend to be significantly different from zero at the .05 level, those higher than .49 at the .01 level, and those higher than .7 at the .001 level. Although levels of significance vary slightly due to changes in number of cases, these figures provide a useful and general rule of thumb for all of the correlations used in this report.

4. Research by Sullivan and Skanes (validity of student evaluations of teaching and the characteristics of successful instructors, Journal of Educational Psychology, 1974, 66 pages, 584-590) has provided evidence of the lack of consistency of the student ratings of relatively inexperienced teachers. Unpublished work at BGSU with the ratings and student test scores of graduate-assistant teachers has also indicated a considerable lack of consistency of graduate-assistant teachers from term to term.

5. Winer, J., Statistical Principles in Experimental Design, Second Edition, McGraw-Hill Book Company, 1971, pages 283-287.

APPENDIX

Page A2 of this appendix is a copy of the letter sent to all departmental faculty to introduce the 1975 peer ranking system. Page A3 presents the criteria for each faculty function to be ranked. Pages A4 to A7 present the student rating instrument and its direction sheet.

The rating procedure used in the Spring of 1974 and the Faculty Evaluation Committee (1975) rating procedures were similar in that all faculty and the five FEC members were provided with a list of all faculty and a name for the three faculty functions (Teaching, Research/Scholarship, Service) and were asked to rate each faculty member on a 1 to 5 scale with "5" representing superior or excellent function and "1" representing poor performance.

Bowling Green State University

Department of Educational
Foundations & Inquiry
Bowling Green, Ohio 43403
(419) 372-0151 ext. 322

A2

April 29, 1975

MEMORANDUM

TO: EDFI Faculty

FROM: EDFI Faculty Evaluation Committee

RE: Peer Input to Faculty Evaluation

After considerable discussion, the EDFI Faculty Evaluation Committee has decided that peer input is an important--and unique--source of information relative to decisions concerning EDFI faculty. The peer input procedure used last year--everyone rate everyone--has too many obvious logical and psychometric disadvantages. The system in which each faculty member asks several peers to provide ratings-recommendations also has many disadvantages. A third approach combines simplicity with psychometric and logical reasonableness--while still producing a type of peer opinion likely to be a valuable supplement to student ratings, chairperson ratings and committee opinion.

Our department is so large and its members' interests and accomplishments are so diverse that it is unreasonable to believe that all of us are aware of the contributions and strengths of all members. However, our faculty are making valuable contributions in the areas of TEACHING and/or RESEARCH and/or SERVICE, and these contributions--many of which are not adequately represented in vita or known to all of us--are known to some of their colleagues. This knowledge can be transmuted into input to FEC decisions via the following colleague-perception-of-contribution system. Each of the following three pages provides a set of questions and/or statements which partially define one of the three areas of academic contribution--Teaching, Research/Production, and Service. Each page also contains a list of EDFI faculty. Faculty members are asked to:

(1) decide upon their own definition of teaching (or research or service);
(2) indicate which faculty member--to their knowledge--best exemplified this definition during the past year;
(3) indicate this person by writing the number "1" in the space next to that person's name;
(4) indicate who is the next best exemplar of their definition by placing a number "2"; and
(5) continue this procedure until a minimum of two faculty and a maximum number of seven faculty are ranked on each of the three areas of contribution (faculty, of course, may nominate themselves in the position that they consider most appropriate).

Faculty resumes for most (many) faculty are available in the departmental office for those who wish to view them. Please return these forms to Cathy Long next week -- May 5 to May 9.

TEACHING

1) Effectiveness in stimulating students to learn
2) Knowledge of content area
3) Effectiveness in sharing teaching competencies with colleagues
4) Efforts to improve teaching effectiveness
5) Effectiveness in advisement
6) Supervision of thesis, dissertation, and/or independent study
7) Development of innovative courses or programs.

SCHOLARLY OR CREATIVE EFFORTS
(Publications, Programs, Research)

1) Has the necessary competencies to produce scholarly or creative efforts
2) Develops proposals, publications, papers, programs, presentations
3) Significant in influence on faculty, organizations, school systems, programs . . .
4) Improves the quality and quantity of scholarly creative efforts by interaction with other faculty, attendance at workshops or professional meetings, extensive reading . . .
5) Functions as a consultant

SERVICE

1) Is an active and valuable contributor to university committees or groups (at the area, department, college, and/or university or state level)
2) Provides service to peers and colleagues
3) Is an active and valuable contributor to professional associations
4) Provides professional public service beyond this campus to assist other universities, colleges, schools, agencies, companies--(not including "good citizenship" activities performed in the capacity of a concerned citizen in church, youth groups, etc.)
5) Has received special awards and/or recognition for professional service.

STUDENT DESCRIPTION OF TEACHING

TO THE TEACHER

Thank you for using the STUDENT DESCRIPTION OF TEACHING. For each class, you will need:

(1) class quantities of the questionnaire;
(2) class quantities of the IBM 1/535 answer sheet, and
(3) one copy of this form to be completed by the teacher of each section.

The procedure for form use involves:

(1) giving the forms, and answer sheets (and some pencils) to a student at the beginning of a class period (BUT NOT DURING THE FINAL EXAM);
(2) writing your name, course name and number, and section number on the blackboard;
(3) leaving the room while one student reads the directions to the class, and the class completes the form;
(4) adding your data sheet (on the back of this page) to the stack of class forms and answer sheets, and asking a student to mail (campus mail) or return (by hand) all forms to: Cathy Long, Dept. EDFI, BGSU (529 Education Building).

We have begun the use of an optically scanned answer sheet in order to avoid the three-week keypunching delay that we faced in the past. Since we will scan—and will be able to use the process computer programs developed last term, we hope to be able to return the results early enough to be of use to you for your next term's classes.

A note about the form:

Page 1. Questions 1 to 5 represent the general teacher qualities most frequently chosen as characteristic of "good teachers" by BGSU students—and by college students in over 40 years of student ratings research. Questions 6-9 are very general questions reflecting the student-perceived effectiveness of the class. Since a summary of such general questions may be the most appropriate general measure of student-perceived classroom teaching success (since it is not reflective of any particular prescribed model of the "ideal" teacher), we define "teacher rating" as the mean score derived from questions 6, 7 and 8.

Some of the other questions on page 1 represent those student and class characteristics which may cause ratings to be biased upward or downward. We are investigating the general problem related to the fairness of comparing large and small classes, undergraduate and graduate classes, etc. A few of these other questions relate to course clarity or difficulty and should be of interest to most faculty.

Back page. The questions on the back page are intended to supply the teacher with more specific information about the class. If you wish to eliminate any (or all) of these questions, do so by including this request in the directions to be read to your class or by not printing the back page when you reproduce the form. If you wish to add your own questions, do so by having your student pass them out with the form —— and modify the directions to indicate this — or by printing them on the back page of the form in place of ours. If you use our questions, please use our numbers.

Please start numbering your own questions—if different from ours—with number 44 and finish with number 70 so that answers to your own questions will not become confused with those of other teachers who use our questions.

We will process and return all data as soon as possible.